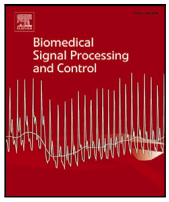
PSC-Net: Integration of Convolutional Neural Networks and transformers for Physiological Signal Classification

Biomedical Signal Processing and Control 91 (2024) 106040

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



PSC-Net: Integration of Convolutional Neural Networks and transformers for Physiological Signal Classification

Qichao Liu [a], Yue Feng [a,*], Hong Xu [b], Jia Li [a], Zhuosheng Lin [a], Shengke Li [a], Shihan Qiu [a], Xin Wu [a], Yuangang Ma [a], Ying Xu [c], Fufeng Li [c,*]

[a] Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, Guangdong, China
[b] Institute for Sustainable Industries and Liveable Cities, Victoria University, Melbourne 8001, Australia
[c] School of Traditional Chinese Medicine, Shanghai University of TCM, Shanghai 201203, China

ARTICLE INFO

Keywords:
CNN
Transformer
WPS
ECG
Physiological signal diagnosis

ABSTRACT

Objective: Physiological signals, such as electrocardiogram (ECG) and wrist pulse signals (WPS), play an important role in diagnosing and preventing cardiovascular and other physiological diseases. Therefore, accurate classification of physiological signals has become the key to assist physicians in diagnosis. However, this field still faces several prominent challenges, including limited availability of data, imbalanced datasets, convergence issues with loss functions, and the need for model architectures capable of accurately detecting waveform patterns.

Methods: This study introduces the Physiological Signal Classification Network (PSC-Net), which combines the strengths of Convolutional Neural Networks (CNNs) and transformers for applications in medical artificial intelligence. Specifically, local temporal features are extracted using the GRWA-LSTM network (GLNet) proposed in this paper. Within the transformer, two GRU layers replace the fully-connected layer to enhance global feature extraction for physiological signal data. Residual connection integrates the outputs of GLNet and Transformer through global average pooling and weight settings. To address challenges related to small and imbalanced datasets, we propose an enhanced data augmentation algorithm based on SMOTE Tomek, along with an improved loss function. Additionally, automatic learning rates are optimized using the Dung Beetle Algorithm (DBA).

Results: Our proposed method achieves superior accuracies of 83.33%, 100.0%, 95.74%, and 98.85% on four physiological signal datasets (including one clinical dataset): Five Types of Pulses Database, Coronary Heart Disease (CHD) database, MIT-BIH Arrhythmia Database, and MIT-BIH ST Change Database. These results attest to the model's robust generalization capability and its promising application prospects in assisting diagnoses.

1. Introduction

According to the latest data from the 2021 European Society of Cardiology Atlas Project, cardiovascular disease (CVD) stands among the leading causes of death. An estimated 17.9 million people die due to CVD annually, constituting 32% of all deaths worldwide [1]. Within these cases, arrhythmia is a prevalent manifestation of CVD, and electrocardiogram (ECG) data plays a crucial role in diagnosing the presence of arrhythmias. Similarly, the wrist pulse signals (WPS) serves as a vital biomarker, reflecting the propagation of physiological phenomena within the arterial system of the human body. By categorizing it into a particular pulse, it can dynamically reflect cardiovascular [2,3]. Moreover, the World Health Organization (WHO) advocates for wrist pulse diagnosis as an efficacious and expedient approach, shown in Fig. 1, for the detection and management of CVD [4], as shown in Fig. 1. ECG and WPS, both physiological signals, are non-invasive and widely utilized medical diagnostic tools. Utilizing deep learning algorithms to classify and process physiological signal data can offer crucial information for medical disease diagnosis and treatment. With the rapid growth of artificial intelligence (AI), deep learning methods have been widely used in many fields including ECG classification. Many methods advocate segmenting the ECG signal into 2, 5 or 10-s heartbeats [5], which are subsequently classified using CNN, Densely Connected CNN (DenseNet) [6], or an improved deep residual network [7]. All these models demonstrate high classification performance. However, in contrast to the ECG signal, the WPS travels through various parts of the body, including nerves, muscles, skin,

* Corresponding authors.
E-mail addresses: j002443@wyu.edu.cn (Y. Feng), li_fufeng@aliyun.com (F. Li).

https://doi.org/10.1016/j.bspc.2024.106040
Received 9 July 2023; Received in revised form 8 December 2023; Accepted 29 January 2024
Available online 3 February 2024

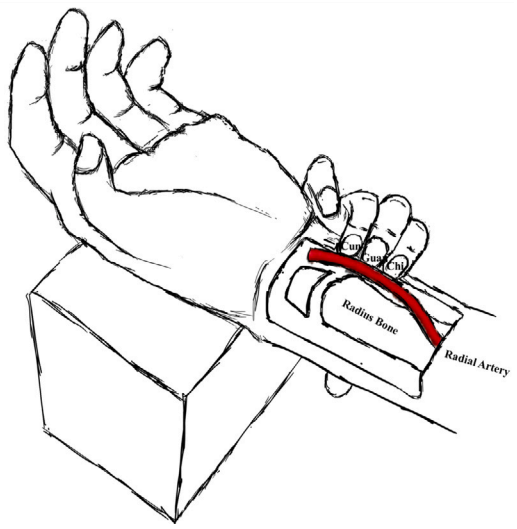


Fig. 1. Wrist pulse diagnosis is typically performed at three specific positions, namely, Cun, Guan and Chi.

and arterial walls. Consequently, the WPS contains more physiological information than the ECG signals [8].

In recent years, there has been a growing interest among researchers in AI-based diagnostic tools for noninvasive physiological signals, with a particular focus on WPS [9–12]. Meanwhile, several challenges have emerged during this exploration process. For instance, Kang et al. [13] utilized the non-invasive monitoring of human WPS to obtain real-time information on heart rate, cardiovascular metrics, and pulse waveforms. Despite utilizing a support vector machine (SVM) for detection, the limited availability of data hindered the capacity of the model to learn diverse patterns and perform effective generalization to new cases, leading to unsatisfactory detection outcomes. Similarly, a wrist pulse acquisition method was proposed [14], generating a feature vector capturing the wrist pulse waveform, and an SVM classifier was employed to distinguish between diabetic and healthy samples. Unfortunately, this dataset suffered from imbalance, with significantly more healthy samples than diabetic samples. This imbalance caused the model to be biased towards the majority category, resulting in suboptimal performance for the minority category. These challenges underscore the need for data enhancement methods to effectively address data scarcity and imbalance in the field of physiological signals. Addressing data imbalance, current research primarily involves two techniques: (1) reducing the majority class size (undersampling [15, 16]) and (2) increasing the minority class size (oversampling [17, 18]). Both undersampling and oversampling can be performed in a stochastic manner with low complexity but may lead to potentially destabilizing behaviors, such as the deletion of important instances or the enhancement of noisy instances.

Furthermore, the prevailing methodology for wrist pulse diagnosis entails manual feature extraction to derive a vector of relevant features, subsequently input into a conventional classifier. Nevertheless, this procedure is labor-intensive and constitutes a substantial impediment to the seamless integration of AI into the field of medicine [19]. To tackle this problem, Wang et al. proposed a 1D convolutional neural network instead of traditional machine learning models for wrist pulse classification [20]. On this basis, researchers [21] proposed a CNN with gate recurrent unit (GRU) and attention architecture to automatically identify and classify gestational periods using wrist pulse data. Yet, these model architectures still have limitations in terms of performance and only apply to specific datasets, which may not satisfy the practical needs encountered in clinical settings.

In this paper, we consider that in real-world scenarios, physiological signals such as ECG and WPS often exhibit multi-periodic patterns, where the variation of each cycle at each time point is not only affected by the temporal pattern of its neighboring regions but also highly correlated with the variation at other time points. This observation naturally leads to a modular structure for time change modeling where the changes in a certain period can be captured in a single module. Moreover, employing pre-processing techniques, such as wavelet transform [22] and trilinear interpolation [23], can aid in decomposing the complex temporal patterns of physiological signals, thereby facilitating the modeling of temporal changes. Accordingly, we combine CNN and Transformer models to build a novel model for ECG and wrist pulse diagnosis using a self-built PSC-Net framework. The CNN part of the model is based on GLNet, which effectively extracts physiological signal features using causal convolution, gated convolution, GRU, and two layers of LSTM. The Transformer block integrates two layers of GRU instead of the traditional fully connected layers, in addition to a residual connection from the global average pooling of the first LSTM layer to the flattened feature, which allows the model to utilize frequency-domain features. To the best of our knowledge, there has been no prior work that employs the fusion of CNN, GRU and Transformer for wrist pulse diagnosis. By appropriately integrating GRU, causal convolution, inflation convolution, and LSTM into the model architecture, it becomes more adaptable to the intra- and inter-cycle variations in physiological signals, thus can effectively model the characteristics of physiological signals and capture more information from the data.

In addition to the improved model architecture, we also propose a new algorithm for data augmentation, which mitigates the impact on accuracy due to data scarcity and imbalance. Our algorithm automatically splits the dataset into training and test sets, applies SMOTE Tomek to the training set to fit minority classes, then add Gaussian noise to the augmented dataset and combines it with the original training set to obtain a larger training set. We also design our loss function by combining focal loss, cross-entropy and triplet loss, and use an algorithm to automatically assign optimal weights to the focal loss, which strategy can be applied to all methods that require optimal weight assignment. Finally, we optimize the loss function using the DBA in the four datasets, which improves the accuracy of the model by approximately 3%.

This study makes the following main contributions:

- We designed an improved data enhancement algorithm based on SMOTE Tomek for the problem of scarcity and category imbalance in clinical datasets, and combined the respective advantages of three loss functions, namely, cross-entropy, focal loss, and triple loss, to optimize the training process using DBA.
- We propose a PSC-Net model specifically for physiological signal data, where we use GLNet to extract local features and a transformer encoder to extract global features. Unlike the existing models with the parallel architecture of the two, our model increases the interaction between the two. We use a residual concatenation with global average pooling to directly connect the output of GLNet to the output of the transformer encoder. This operation not only eliminates a large amount of redundant information in the input of the transformer encoder, but also allows the model to retain previously extracted local features. The inter-period and intra-period information is effectively captured and performs well in all four datasets.
- This work introduces an automated weight assignment algorithm for focal loss, effectively addressing the challenge of setting appropriate weights across diverse scenarios.

2. Related work

The modeling and analysis of physiological signals for medical disease diagnosis have been fundamental applications in the field of time series data and have received extensive research attention. Addressing issues such as data imbalance, the extraction of better feature

vectors, and the optimization of the model training process are urgent challenges. In this section, we categorize related work into three main areas: the synthetic minority oversampling technique (SMOTE) methods, model construction, and model optimization.

2.1. SMOTE methods

While considerable research has been conducted on addressing class imbalance challenges in shallow machine learning models, limited attention has been devoted to tackling this issue in the realm of deep learning. Historically, two primary avenues have been pursued to mitigate this challenge: modifications to loss functions and the adoption of resampling techniques. The primary method for modifying the loss function scheme involves using the focal loss function. However, this type of function is considered to require artificial setting of weight parameters, introducing significant uncertainty. Deep learning resampling solutions can be broadly categorized as either pixel-based or reliant on generative adversarial networks (GANs) for the generation of synthetic instances. However, both of these approaches come with inherent limitations.

Pixel-based resampling solutions often fall short in capturing the intricate data characteristics of images and may struggle to produce semantically meaningful artificial images [24]. On the other hand, GAN-based solutions demand substantial amounts of data, are challenging to fine-tune, and are susceptible to the problem of mode collapse [25,26]. Therefore, There is a pressing need for a new oversampling method is needed that specifically addresses the challenges posed by the small amount of data and the extreme imbalance between categories in deep learning.

SMOTE has gained significant popularity in various fields, including machine learning classification problems, image recognition and bioinformatics. It was introduced by Megahed et al. [27], with the aim to resolve the issue of imbalanced datasets by generating synthetic minority samples through the interpolation of existing samples. The authors provided a detailed explanation of the principles of their algorithm, conducted multiple sets of experiments to validate its effectiveness.

The SMOTE Tomek algorithm selected in this paper offers the advantage of integrating oversampling and undersampling techniques in a single step, unlike other SMOTE algorithms that generate only a few synthetic samples to address the class imbalance. Specifically, it utilizes the SMOTE algorithm to identify boundary instances and subsequently employs the Tomek link undersampling technique to remove them. This comprehensive approach effectively addresses the challenges associated with unbalanced class distribution and overlapping instances near decision boundaries. Thus, the utilization of the SMOTE Tomek algorithm brings the potential to enhance the classification performance by improving the quality and balance of the dataset.

However, SMOTE Tomek also has some limitations, specifically the lack of diversity in the data generated. This limitation can result in the model to being well-trained on only one aspect of the data's features during subsequent training, inevitably impacting the final accuracy. In this paper, we address this limitation by enhancing the diversity of the data through additional processes, such as noise addition and Fisher–Yates disruption, building upon the foundation of based on SMOTE Tomek. The detailed content is explained in the subsection of Data augmentation algorithm in the Section 3.2, along with the corresponding algorithm implementation process, as outlined in Algorithm 1.

2.2. Model construction

In recent years, many popular deep learning models have been proposed, such as ConvNeXt [28], ViT (Vision Transformer) [29], Swin Transformer [30], and more. Technically, ConvNeXt is based on convolutional neural network that utilizes convolutional layers to capture image features. ViT employs a self-attention mechanism, dividing the image into blocks and transforming these blocks into sequences for image classification tasks. Swin Transformer combines local windowing and global windowing with multilevel attention mechanisms to effectively handle both local and global information, demonstrating strong performance in image classification and other visual tasks. It is noteworthy that despite the success of these popular models in image processing, none of these approaches consider the specific characteristics of time series data, i.e., the trending and periodicity inherent in one-dimensional data. This paper aims to address this gap and focuses on incorporating these characteristics into the modeling of time series data.

We observe that Transformers excel at detecting long-time dependencies between different time points through attention mechanisms. For instance, Wu et al. proposed Autoformer, integrating an autocorrelation mechanism to capture serial temporal dependencies based on learning cycles [31]. However, recognizing only long-term features may be insufficient. This paper recognizes the need for additional local features and introduces a CNN architecture for this purpose. Unlike many existing models that simply divided into two paths to extract local and global features independently and then fuse them in the final fully connected layer, such an approach is deemed unscientific. It often leads to the model relearning features from scratch each time, with limited mutual benefit. In contrast to previous methods, the PSC-Net introduced in this paper transcends the parallel architecture of CNN and Transformer, fostering more interaction between them. The proposed model first extracts local features using the GLNet architecture, an improved version of CNN. It then splits into two paths: one is directly connected to the transformer encoder architecture. In this configuration, the GLNet acts as a facilitator for the transformer encoder, conduction feature filtering to eliminate redundant information. The other path involves taking the local features filtering extracted from the GLNet architecture, pooling them through global averaging to compress the space and leverage their ability to extract low-frequency frequency-domain features. These features are then weighted and fused with the output of the transformer encoder, with weights of 0.64 and 0.35, respectively. Finally, the 4-layer MLP completes the final category output. is completed. It is worth mentioning that our model architecture, tailored for time series data the characteristics, incorporates modules like GRU and LSTM, which are particularly effective for modeling physiological signals.

2.3. Model optimization

In the realm of model optimization, contemporary methods often rely on a single loss function, such as cross entropy or focal loss. Cross-entropy, a traditional loss function commonly used in classification tasks, faces challenges in handling category imbalance and overlook minority classes in multi-category classification scenarios. On other hand, focal loss addresses the issue of category imbalance by enhancing the weight of difficult-to-categorize samples through the focus parameter. While focal loss is effective, its reliance on manual setting of the focus parameter introduces uncertainty. The triplet loss, frequently employed in contrast learning, is instrumental in learning feature embedding space. However, its direct applicability to multi-category classification tasks can be challenging. In this paper, we aim to overcome the limitation of each individual loss function by combining cross-entropy, focal loss, and triplet loss.

Cross-entropy loss aids in the classification problem, focus loss tackles imbalance, and triplet loss contributes to feature embedding. Additionally, to address the issue of setting weight for the focal loss function, this paper introduces an automatic weight setting algorithm. This algorithm adjusts weight based on the number of categories in the input data, assigning greater weights to samples within underrepresented categories. The targeted training approach is elaborated in Section 4.4.

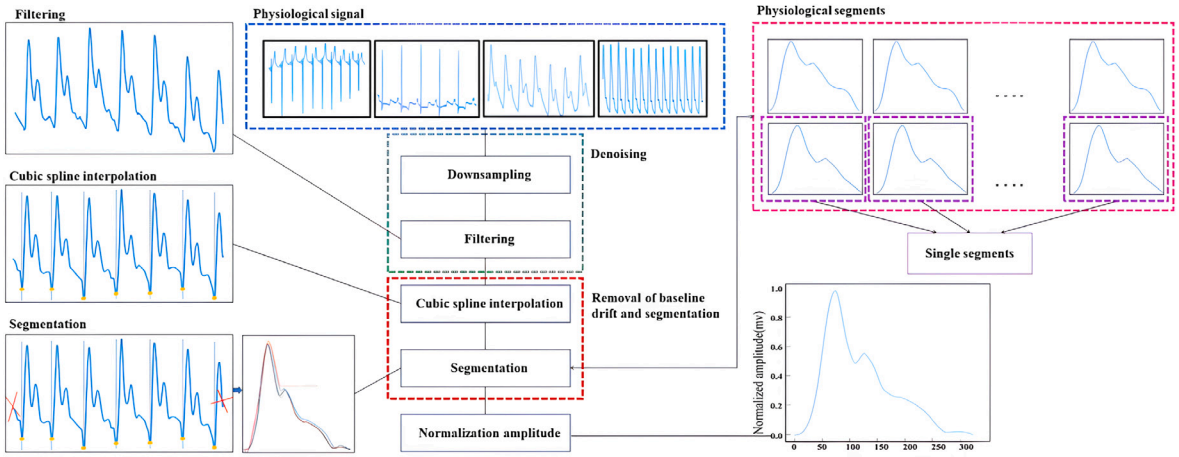


Fig. 2. A preprocessing operation is performed on the physiological signals to obtain single-cycle signals for physiological signal anomaly identification.

Furthermore, our choice to employ the DBA over the current annealed cosine algorithm stems from the inherently multi-periodic nature of the input data. The DBA excels in optimizing multi-peak functions, employing both global and local search strategies harmoniously. This dual strategy proves valuable in avoiding the pitfalls of local optima. This efficacy of the DBA is evident in the experimental results presented in Model 12 and Model 13 within Table 14, highlighting the performance enhancements achieved by incorporating the DBA in this study.

3. Datasets

We validate the model's disease diagnosis performance using biosignals with four datasets, including two private and two public datasets. In addition, clinical medical data require specific pre-processing due to high and low-frequency noise interference. The denoised data better align with real-life data patterns, enhancing the model's suitability for disease diagnosis. The data processing flow is shown in Fig. 2.

3.1. Data description

(1) Five Types of Pulses Database: We recruited a cohort of volunteers from Wuyi University, with data collection conducted by a team of five research assistants who received rigorous training by physicians to ensure the highest level of experimental reliability, and performed the entire study in ethical compliance with the guidelines of the Human Research Ethics Committee (approval number: [2019] 18). To capture the wrist pulse waveform data, we utilized a state-of-the-art lingual pulse and meridian point information collection management system (model MT-SM-01), which has been awarded the prestigious Class II medical device registration certificate by Tianjin Huimin Technology. The collected data were meticulously organized and presented in the form of organized .xlsx tables. Five types of wrist pulse waves (or Mais) were included in the dataset, namely, hesitant pulse, slippery pulse, sluggish pulse, moderate pulse, and counted pulse.

(2) CHD Database: In this work, we also included pulse data from patients with CHD, which were generously provided by the Shanghai University of Traditional Chinese Medicine. This dataset consists of two distinct patient categories: individuals with a health condition and those diagnosed with coronary artery disease.

Tables 1 and 2 comprehensively present the number of beats per category for the Five Types of Pulses Database and two types of CHD Database, respectively, accompanied by detailed explanations of their clinical significance.

(3) MIT-BIH Arrhythmia Database [32]: The single-lead database comprises 48 half-hour ECG recordings obtained from the physionet database. Featuring a sampling frequency of 360 Hz and 11-bit resolution, this database includes a rich dataset for developing and evaluating algorithms for arrhythmia detection and classification. The source data is available at https://www.physionet.org/content/mitdb/1.0.0/

(4) MIT-BIH ST Change Database [33]: This single-lead database focuses on ST segment changes that serve as crucial indicators of cardiac conditions, particularly ischemia. It consists of 90 annotated ECG recordings from 59 patients, each lasting 30 min. The source data is available at https://www.physionet.org/content/stdb/1.0.0/

The MIT-BIH Arrhythmia Database was used in a 7-class classification task, including 'Normal', 'Left bundle branch block', 'Right bundle branch block', 'Premature ventricular contraction', 'Atrial premature beat', 'Fusion of ventricular and normal beat', and 'Fusion of paced and normal beat'. On the other hand, the MIT-BIH ST Change Database was used in a 3-class classification task, including the classes of 'Normal', 'ST Elevation', and 'ST Depression'. The specific data distributions are shown in Tables 3 and 4.

Table 1
Number of beats in each class of pulse and explanation of Five Types of Pulses Database.

Data type	Clinical relevance	Grouped class
Hesitant pulse	Hesitant pulse (Chi Mai), characterized by a slow or irregular rhythm of the pulse, is a common symptom of cardiovascular diseases.	99
Slippery pulse	Slippery pulse (Hua Mai) may be caused by phlegm, accumulated food, or actual heat.	444
Sluggish pulse	Sluggish pulse (Huan Mai) is mostly caused by spleen deficiency and lack of qi and blood.	407
Moderate pulse	In case of moderate pulse (Ping Mai), the body is in a state of harmony between yin and yang.	1317
Counted pulse	Counted pulses (Shu Mai) are most often seen in heat evidence, strong for actual heat, weak for deficiency heat.	609

Table 2
Number of beats in each class of pulse and explanation of CHD Database.

Data type	Clinical relevance	Grouped class
CHD	CHD is a condition characterized by a narrowing or blockage of the coronary arteries, which can lead to chest pain, shortness of breath, and even heart attack.	1270
Normal	Normal individuals do not have such blockages and do not exhibit symptoms of CHD.	900

Table 3
Types of QRS beat in the MIT-BIH Arrhythmia Database.

Symbol	Heartbeat type	Grouped class
N	Normal beat	5921
L	Left bundle branch block beat	949
R	Right bundle branch block beat	977
A	Atrial premature beat	540
V	Premature ventricular contraction	2162
F	Fusion of ventricular and normal beat	217
f	Fusion of paced and normal beat	43

Table 4
Types of QRS beat in the MIT-BIH ST Change Database.

Symbol	Heartbeat type	Grouped class
N	Normal beat	6369
S	ST Elevation	478
V	ST Depression	111

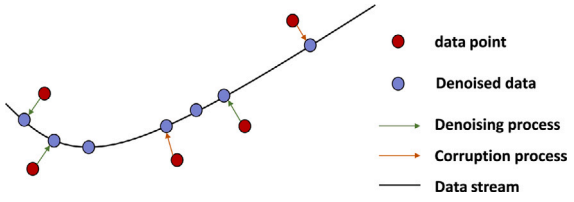


Fig. 3. Denoising and interpolation of data.

3.2. Data preprocessing

Denoising. During physiological signal acquisition, minute body tremors, breathing and power frequency interference from instruments can cause high- and low-frequency noise and baseline drift. Studies have indicated that noise induced by power frequency interference and circuit interface typically appears above 20 Hz, while external interference, such as breathing and arm movements, usually occurs below 0.1 Hz [34,35]. Bandpass filters remove both high- and low-frequency noise [36]. Discrete cosine transform, empirical mode decomposition and wavelet transform can eliminate baseline drift [37]. Moreover, cubic spline interpolation provides a way to smooth denoised WPS [38].

In this study, we first used a Balwons bandpass filter to eliminate noise outside the 0.5 Hz to 40 Hz range as a means to exclude 50 Hz industrial frequency interference and low-frequency noise below 0.5 Hz. Subsequently, we performed a 5-layer decomposition of the WPS using db6 wavelet basis in the wavelet transform. Since the sampling frequency of the WPS data set is 128 Hz, the frequency range of the signal is 0~64 Hz, and the bandwidth of the high-frequency detail components D1 and D2 is 64 Hz~16 Hz. Considering that most of the energy of the human WPS is mainly concentrated to between 0~10 Hz, we reconstructed the high-frequency detail components of D1 and D2 to complete the high-frequency noise of the WPS removal and retain most of the energy of the WPS. In addition, the other datasets were processed similarly to the WPS dataset as shown in Fig. 3.

Removal of baseline drift and segmentation. Since different periods will lead to intro period and interpreted variations, it is straightforward to think that by splitting the WPS into multiple single cycles and using the designed network architecture, we can capture the cycle-specific derived variations in a single module and thus obtain the intra- and intercycle characteristics. Besides, this design makes intricate temporal patterns disentangled, benefiting oral variation modeling. This paper utilizes cubic spline interpolation to perform WPS segmentation and baseline drift removal (as shown in Fig. 3). The process involves calculating the fundamental frequency of the WPS to determine the length of the single-cycle WPS. Then, a corresponding sample point window is designed, and the window is moved backward until a complete WPS is scanned. The window is then interpolated three times to achieve curve fitting, and the first-order derivative is used to obtain the corresponding minimum value, which determines the minimum value of the window. Next, three spline interpolations are performed to fit the curve to the entire wrist pulse signal and each minimum. The baseline drift is removed by subtracting the corresponding minimum curve from the original wrist pulse signal curve. Finally, the list of recorded minima can be segmented to obtain the corresponding single-cycle pulse waveform, exactly as described in Fig. 2.

Algorithm 1 Data augmentation algorithm

Input: Dataset X consisting of n samples with d features each.
Split ratio for training and test sets, λ_1.
Split ratio for training and validation sets, λ_2.
Number of nearest neighbors for SMOTE algorithm, k.
Amount of Gaussian noise to be added, σ.
Output: Augmented dataset $\{X\}_{train}$.

Split the dataset X into $\{X'\}_{train}$ and $\{X'\}_{test}$ using the split ratio λ_1
Apply SMOTE Tomek to the training set
for $\forall \quad X_i \in \{X'\}_{train}$ **do**
  Calculate distances the k nearest neighbors of X_i
  for each nearest neighbor X_i' **do**
    Generate a synthetic sample by interpolating X_i and X_i'
  end for
end for
Get the characteristics of $\{X''\}_{train}$
Add Gaussian noise σ
if $\forall\, X_i \in \{X''\}_{train}$ **then**
  for $i = 1$ to n **do**
    $X_i + \sigma = X_i'$
  end for
end if
Get the characteristics of $\{X'''\}_{train}$
Combine the $\{X'''\}_{train}$ with the $\{X''\}_{train}$ to create the augmented training set, then using the Fisher-Yates algorithm to disrupt
Split the augmented training set using the split ratio λ_2
return $\{X\}_{train}$.

Data augmentation algorithm. The uniform distribution of data between training and test sets is crucial to achieve high classification accuracy. To this end, we utilize the Fisher–Yates algorithm for shuffling the arrays, as it offers distinct advantages over conventional shuffling methods. Specifically, the Fisher–Yates algorithm ensures that the position of each element in the array is equally likely to be scrambled, resulting in a more balanced distribution. Additionally, the time complexity of the algorithm is linear, i.e., O(n), where n denotes the length of the array. Furthermore, its space complexity is constant, i.e., O(1), thus it requires a minimal number of auxiliary variables to swap array elements. It is evident that the distribution of data is highly imbalanced, as indicated in Fig. 4(a), with a ratio of approximately 1:13.17 for the category 'Hesitant pulse' to the category of 'Moderate pulse'. Such data imbalance is a common challenge encountered in clinical medical data analysis. To address this issue, this paper proposes a novel algorithm specifically designed to handle imbalanced physiological signal datasets. This algorithm integrates several preprocessing methods, including adding Gauss Noise, rotation, translation, Borderline SMOTE, and SMOTE Tomek, to make up for the lack of data preprocessing methods in the field of physiological signals and make it easy for later generations to call it directly. In our experiment, the algorithm takes the segmented training set as input and performs a Smote fitting few-class operation to obtain a balanced training set, as shown in Fig. 4(b).

However, since some of the fitted features may not conform to the principle of feature diversity, the algorithm replicates the training set and adds Gaussian noise to enhance diversity. The resulting dataset is then merged with the original training set. Finally, the results are randomized using the Fisher–Yates algorithm to form the final training set

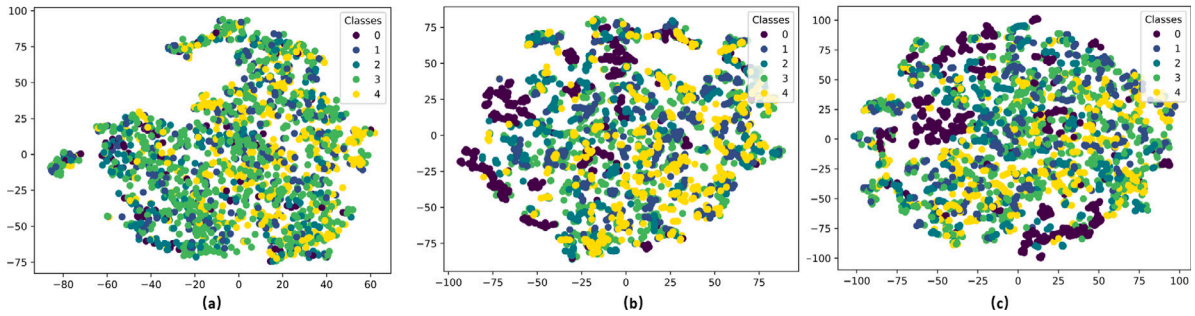


Fig. 4. Using t-SNE to visualize the data augmentation process, taking Five types of pulse database as an example: (a) using Fisher–Yates algorithm (b) using the SMOTE Tomek algorithm (c) using data augmentation algorithm results.

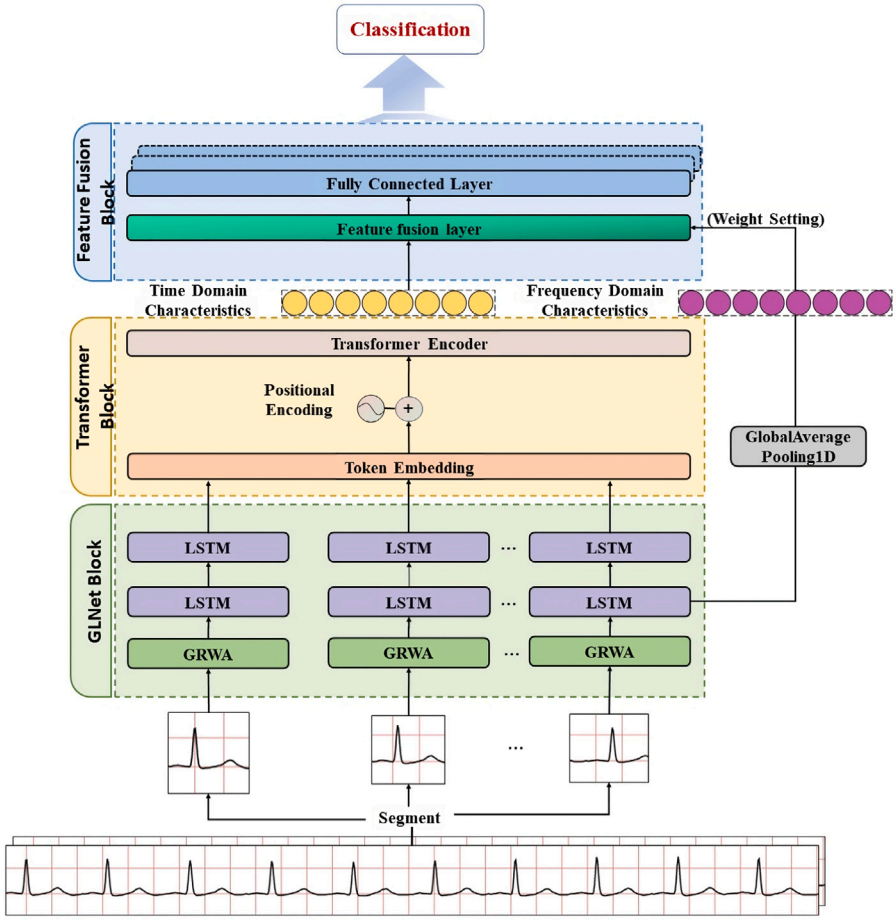


Fig. 5. Overall architecture of PSC-Net, which consists of three components: the GLNet Block can capture local features; the Transformer block can extract global features and establishes correlations between the signals; the feature fusion block does the final sorting.

for model training. The detailed procedure is illustrated in Algorithm 1, and the corresponding outcomes are visualized in Fig. 4(c).

4. Methods

This section presents the PSC-Net, a novel architecture for physiological classification that combines both local and global feature extraction. As illustrated in Fig. 5, the proposed architecture consists of three components: (1) The GLNet Block, which is specifically designed to effectively capture the local features of biological signals; (2) A modified transformer encoder block that extracts global features and establishes correlations between the signals, adapted to the characteristics of physiological signal data to enhance the ability of the model to learn complex temporal relationships; (3) Finally, the feature fusion block that integrates multilevel features from previous layers to produce the final categorical output. A detailed description of these three components is provided in the following three sections. Our proposed model is trained on a combination of ECG and pulse data, collectively referred to as biological signals, resulting in a highly robust and versatile framework for processing and analyzing complex biosignals.

4.1. GLNet block

GRWA Block. The overall GRWA model architecture is shown in Fig. 6. Given an ECG time or Pulse series signal, this module seeks to

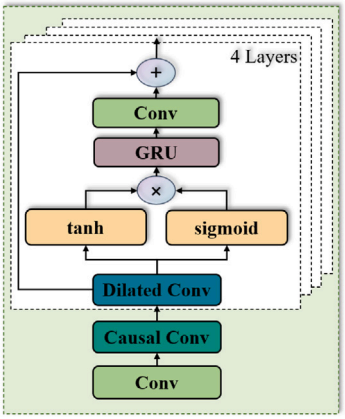


Fig. 6. Structure of GRWA.

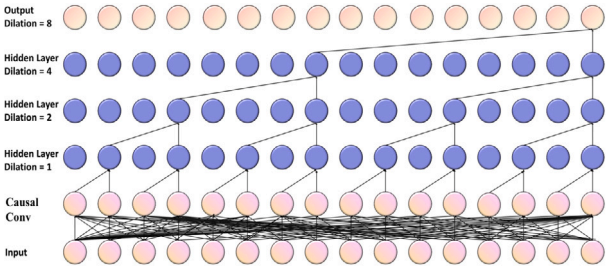


Fig. 7. Part of the GRWA network architecture which includes standard convolution, causal convolution, and expansion convolution.

extract spatiotemporal information. Let X be a physiological signal that has been recorded and separated into n segments, i.e., $X = \{x_i\}_{i=1}^{i-n}$. Due to the requirement that the input of the neural network has a fixed length, l is chosen as the length of each segment. The proposed approach in this work leverages a combination of causal convolution, dilated convolution and gated convolution. Causal convolution is employed to account for the causal relationship of the input in the temporal domain, dilated convolution is utilized to capture a larger receptive field and extract more features (as shown in Fig. 7), and gated convolution is incorporated to regulate the flow of information through gating mechanisms, enabling the network to effectively capture long-term dependencies in the input signal. Meanwhile, GRU, as a variant of RNN, excels in addressing long-term memory and gradients in backpropagation. In addition to the above, we also incorporate a standard convolution layer, which helps to extract low-level features. The integration of causal convolution, dilated convolution, gated convolution, GRU, and standard convolution enables our approach to capture short- and long-term dependencies in physiological signal data in a highly effective manner. This synergistic combination effectively compensates for the weaknesses inherent to each individual operation, resulting in a more robust and powerful model that can effectively handle complex physiological signal data. We represent 1D CNNs that use the function F.

$$\{f_i\}_{i=1}^{i=n} = F\left(\{x_i\}_{i=1}^{i=n}\right). \tag{1}$$

LSTM Block. A LSTM [39] is applied to model the temporal relations between 1D segments. Let us represent the output of the LSTM and embody its temporal modeling function as:

$$\{g_i\}_{i=1}^{i=n} = S\left(\{f_i\}_{i=1}^{i=n}\right). \tag{2}$$

with $f_i \in S^L$. Furthermore, the output of this module $\{g_i\}_{i=1}^{i=n}$ is subsequently sent to the transformer encoder module.

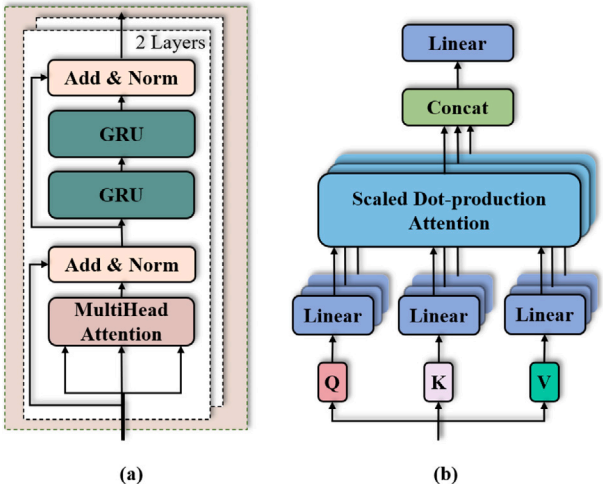


Fig. 8. (a) Structure of transformer block (b) Structure of multi-head attention.

4.2. Transformer block

In our proposed module, the LSTM Block outputs are fused by 2 layers of transformer encoder, which aids the model in avoiding missing information. The entire network is illustrated in Fig. 8. In the first step, the input is reshaped and subjected to positional encoding. At the second step, the transformer encoder [40] is taken to reweight these features by a proper ratio. These elements have dimensions of $N \times d$, where N represents the maximum sequence length and d is the length of the feature array. This process is illustrated in Fig. 8(a), where the input is passed through three fully connected layers. Considering the difference between physiological signal data and other types of data such as images, we choose to use two GRU layers instead of the original fully connected layer. This modification has been proved as effective in experiments, and the average accuracy is improved by 4%. The final feature T value is computed as follows:

$$T = \{\mathrm{w_i}\}\{g_i\}_{i=1}^{i=n}. \tag{3}$$

where the values of w_i are learned by the transformer encoder.

It is worth mentioning that the self-attention mechanism takes the input series and transforms it into three elements: query (Q), key (K) and value (V), as described in Eqs. (4)–(6). These elements have dimensions of $N \times d$, where N represents the maximum sequence length and d is the length of the feature array. The process is illustrated in Fig. 8(b), where the input is passed through three fully connected layers. Then, the self-attention mechanism performs a transformation on the query and the set of key–value pairs to generate an output, as indicated in Eqs. (7)–(8). The output of the self-attention mechanism, denoted as O, is obtained by multiplying the value (V) with the attention matrix B.

$$Q^i = (q_1, q_2, \ldots, q_n), \tag{4}$$

$$K^i = (k_1, k_2, \ldots, k_n), \tag{5}$$

$$V^i = (v_1, v_2, \ldots, v_n), \tag{6}$$

$$B = \mathrm{Softmax}\left(\frac{QK^i}{\sqrt{d}}\right), \tag{7}$$

$$O = BV. \tag{8}$$

4.3. Feature fusion block

In this section, we introduce a novel feature fusion method that differs from the traditional fully connected layer. Our approach fuses

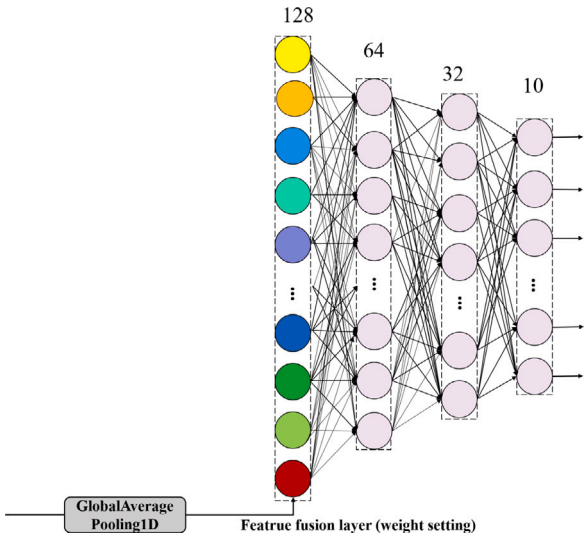


Fig. 9. Architecture of feature fusion block.

the features extracted from the transformer encoder with the LSTM output via residual concatenation. This enables the model to capture long-time feature dependencies using the transformer encoder without overlooking previously extracted local features. Simultaneously, the GLNet architecture effectively aids the transformer encoder in eliminating irrelevant redundant information, as illustrated in Fig. 9.

Within the feature fusion layer, morphological features extracted from the transformer encoder and frequency domain features obtained by Global Average Pooling are combined. This enables the model to retain both the locally and globally extracted features. We assigned weights of 0.65 and 0.35 to the features generated by GLNet and the features produced by the transformer encoder, respectively. Subsequently, the final classification results are obtained after passing through fully connected layers with parameter settings of 128, 64, 32, and 10, respectively.

To further optimize the feature fusion process, we introduced global average pooling to the residual concatenation to compress the spatial size of each channel. Additionally, leveraging the Discrete Cosine Transform (DCT) property of global average pooling, which can extract the lowest frequency, proves beneficial for feature fusion. Furthermore, inspired by the concept of BiFPN [41], we assigned a weight bias to the fusion to enhance its focus. We experimented with various weight settings, including (0.2, 0.8), (0.3, 0.7), and (0.4, 0.6). Ultimately, we found that setting the weights to (0.65, 0.35) consistently yielded the best results across all four models. This preference may be attributed to the fact that the designed GLNet architecture captures more critical local information that is more favorable to the model.

4.4. Automatic weight setting

In order to improve the performance of focal loss, it is usually necessary to set the corresponding weight parameters for each category, significantly influencing the final performance. However, human intervention in this process introduces substantial uncertainty. To address this challenge, this paper designs an algorithm for automatic weight setting. The core idea of the algorithm is to assign greater weights to categories with fewer instances, allowing the model to better learn this underrepresented aspect, a crucial consideration often overlooked in contemporary models. To achieve this task, the algorithm first calculates the weight of each category W_c by applying the formula:

$$W_c = \frac{C_s}{\mathrm{Num}_c \times C_c} \,. \tag{9}$$

where C_s represents the number of samples in the training set, Num_c represents the number of categories of labels, and C_c is the number of occurrences of each category in the training set. The resulting weight for each category is then returned as an array, which can be used during the training process to ensure that the model pays more attention to under-represented classes. The corresponding pseudocode implementation is shown in Algorithm 2.

Algorithm 2 Automatic weight setting

Input: Data for which weight settings are required Y_{train}.
Output: Parameters of class weights in focal loss $class\_weights$.

```
def get_class_weights(Y_train):
    num_samples = Compute total number of samples in Y_train
    num_classes = Get the total number of classes in Y_train
    class_counts = Sum of each class count in Y_train along axis 0
    class_weights = num_samples / (num_classes × class_counts)
    return class_weights
class_weights= get_class_weights(Y_train)
```

5. Results

5.1. Experimental settings

In the present study, two confidential pulse datasets were employed along with the MIT-BIH Arrhythmia Database and the MIT-BIH ST Change Database retrieved from the eminent Physionet public database, yielding in a total of four datasets that were utilized to analyze and evaluate the performance of the proposed algorithm. For the experiment, we partitioned the complete dataset into training, validation and test sets at a ratio of 6:2:2. To maintain data balance, we extracted samples for each set from sub-datasets of various categories proportionally to their contribution to the overall dataset. This approach ensured that each category was adequately represented in the training, validation and test sets. The model was trained for 120 epochs with a batch size of 16. To facilitate experimentation, we employed resampling techniques on our datasets that are similar to earlier work by J. Zhang et al. [42]. The ECG datasets were uniformly downsampled to 328 and the WPS was downsampled to 320. G. AlMahadin et al. [43] claimed that this resampling approach enhances the reliability and reproducibility of experimental results as it allows for more consistent and standardized data preprocessing across all datasets used in our study. Other data preprocessing operations are explained in detail in Section 3 Summary Data Preprocessing.

We used an end-to-end approach to train the model using the Nadam optimizer. Our approach utilizes a novel loss function that combines the weighted fusion of cross-entropy, focal loss and triplet loss. The cross-entropy loss function can determine how close the actual output is to the desired output and is represented by the following formula:

$$L_{ce} = \frac{1}{N}\sum_i L_i = -\frac{1}{N}\sum_i \sum_{x=1}^{C} y_{ix} \log\left(K_{ix}\right) \tag{10}$$

where C represents the number of categories; $\mathbf{y_{ix}}$ is the indicator function (0 or 1) that takes the value 1 if the true category of sample i is equal to x, otherwise it takes 0; K_{ix} denotes the predicted probability that sample i belongs to category x.

The focal loss effectively addresses data imbalance issues, while triplet loss encourages the model to learn more discriminative feature representations by comparing within-category sample similarities and between-category differences. The equations for the two loss functions are presented as follows:

$$L_{FL}(\mathbf{K_t},) = -alpha(1-\mathbf{K_t})^{\gamma} log(\mathbf{K_t})\,, \tag{11}$$

$$K_t = \begin{cases} K & \text{if y = 1} \\ 1-K & \text{otherwise} \end{cases}, \tag{12}$$

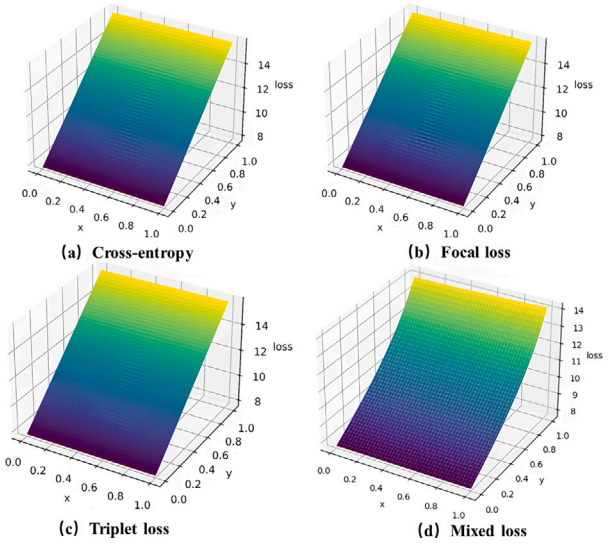


Fig. 10. (a), (b), and (c) illustrate the loss landscapes of different loss functions: cross-entropy, focal loss and triplet loss, respectively. These plots show relatively flat loss curves without prominent concave regions. However, Fig. 10(d) demonstrates a distinct character with a smooth and concave-shaped loss landscape, indicating that the proposed weighted fusion loss function leads to better convergence and achieves lower loss values.

$$L_{tr} = \max(0, \|f(an) - f(\text{pos})\|^2 - \|f(an) - f(\text{neg})\|^2 + mar). \tag{13}$$

The balance coefficient $alpha$ is determined using an automatic weight-giving algorithm developed by the author of this study. The degree of focus, γ, which controls the emphasis on difficult samples, is set to 2. The predicted positive class probability K_t is used in the calculation. In Eq. (12), we denote $f(an)$ as the feature representation of the anchor sample, $f(pos)$ as the feature representation of the positive sample, and $f(neg)$ as the feature representation of the negative sample. The Euclidean distance between features is represented by $\|\cdot\|$. We introduce the term mar, which is a predefined boundary value aimed at controlling the difference in distance between positive and negative samples. The above formulation allows us to effectively compute the triplet loss, ensuring that the positive sample is closer to the anchor sample than the negative sample by a margin of at least mar. Optimizing this loss function can encourage the model to learn discriminative feature representations for different classes in the context of multi-class classification tasks. To integrate these three loss functions, we employ a weighted fusion approach. Specifically, we assign weights of 0.3, 0.5, and 0.2 to each respective loss function. By considering these weights, we formulate the final fusion as follows:

$$FinalLoss = 0.3L_{ce} + 0.5L_{FL} + 0.2L_{tr} \tag{14}$$

Fig. 10 illustrates the smoother landscape achieved by our composite loss function, which results in faster convergence and higher accuracy compared to a single loss function, as further supported by the ablation experiments detailed in Table 14. We adopted an initial learning rate of 0.0001 in our training process, which was determined via experimentation. In addition, we introduced the DBA to dynamically adjust the learning rate during training, which resulted in further optimization. Our approach was compared against other optimization techniques, such as the annealed cosine algorithm, the chimpanzee algorithm, and the no-optimization algorithm.

The proposed approach is aimed at using the python language; the PyTorch demonstrates the effectiveness of our approach in achieving superior model accuracy and faster convergence rates than the other techniques 1.12.1 framework. All experiments in this paper were run on a computer with an Intel(R) Core(TM) i7-11700K CPU environment and NVIDIA GeForce GTX A5000 GPU acceleration environment, using an Ubuntu OS with 24 GB memory.

Table 5
Performance metrics for identifying Five Types of Pulses Database. The best results are shown in bold (see [44–50]).

Method	ACC (%)	PR (%)	RE (%)	F1 (%)
RF [3]	65.97	59.15	43.58	46.81
SVM [44]	65.45	55.40	52.70	53.05
XGBoost [45]	70.13	57.07	63.95	59.51
LSTM [3]	79.86	74.22	68.38	69.83
GRU [46]	80.21	74.35	72.11	72.83
CNN+LSTM [47]	74.31	68.77	70.73	69.50
CNN [20]	57.99	49.75	58.03	51.36
ResNet34 [48]	70.49	61.84	62.37	62.09
ResNet32+se	71.18	62.17	64.18	62.99
ResNet32+se+LSTM	75.00	63.37	68.83	63.92
ViT [29]	75.34	67.89	64.30	65.85
ConvNeXt-S [28]	77.77	64.79	63.86	64.05
CNN-SVM [49]	79.53	68.24	70.58	69.72
CNN + BiLSTM [50]	81.46	76.57	73.92	73.88
PSC-Net (ours)	**83.33**	**79.93**	**75.52**	**76.99**

Table 6
Performance metrics for identifying two types of CHD Database. The best results are shown in bold.

Method	ACC (%)	PR (%)	RE (%)	F1 (%)
RF [3]	99.05	99.10	98.97	99.04
SVM [44]	99.53	99.50	99.50	99.50
XGBoost [45]	99.53	99.58	99.46	99.52
LSTM [3]	98.58	98.46	98.62	98.54
GRU [46]	99.05	98.89	99.19	99.03
CNN+LSTM [47]	98.58	98.39	98.61	98.50
CNN [20]	98.58	98.61	98.52	98.57
ResNet32 [48]	99.53	99.60	99.43	99.51
ResNet32+se	99.53	99.57	99.47	99.52
ResNet32+se+LSTM	99.53	99.63	99.36	99.49
ViT [29]	98.29	79.15	90.00	86.83
ConvNeXt-S [28]	99.05	99.20	98.86	99.02
CNN-SVM [49]	99.11	98.02	99.23	99.24
CNN + BiLSTM [50]	99.80	99.46	99.01	99.57
PSC-Net (ours)	**100.0**	**100.0**	**100.0**	**100.0**

5.2. Evaluation metric

In order to quantify the performance of our model, we employ the commonly used evaluation metrics of accuracy (ACC), precision (PR), recall (RE), and F1-score (F1). The detailed calculation formulas are described below.

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \tag{15}$$

$$PR = \frac{TP}{TP + FP} \times 100\% \tag{16}$$

$$RE = \frac{TP}{TP + FN} \times 100\% \tag{17}$$

$$F1 = \frac{2 \times PR \times RE}{PR + RE} \times 100\% \tag{18}$$

where FP, FN, TP, and TN represent false positive, false negative, true positive, and true negative, respectively. These variables are used to calculate the evaluation metrics, which can measure the classification performance of the model.

5.3. Experimental results

In this study, we utilize 5-fold cross-validation to ensure the reliability of results. The final result is obtained by averaging the results of 5 tests. The data demonstrate that our proposed model achieves higher

Table 7
Performance metrics for identifying seven types of MIT-BIH Arrhythmia Database. The best results are shown in bold (see [51–55]).

Method	ACC (%)	PR (%)	RE (%)	F1 (%)
CNN [51]	89.82	81.88	67.98	71.30
ECVT-Net [52]	92.87	79.04	74.83	76.54
ECT-net [53]	89.34	79.45	72.18	75.19
ConvNeXt-S [28]	95.28	82.85	81.11	81.74
CNN+Transformer [54]	92.60	89.04	79.43	83.36
CNN+LSTM [47]	94.73	85.96	85.17	84.92
CNN+Transformer [55]	86.03	65.45	59.81	61.86
ViT [29]	86.12	61.69	58.42	59.63
CNN-SVM [49]	94.40	83.70	82.30	83.52
CNN + BiLSTM [50]	94.93	88.24	**85.68**	85.11
PSC-Net (ours)	**95.74**	**89.93**	84.06	**85.85**

Table 8
Performance metrics for identifying three types of MIT-BIH ST Change Database. The best results are shown in bold.

Method	ACC (%)	PR (%)	RE (%)	F1 (%)
CNN [51]	95.83	90.24	74.19	80.78
ECVT-Net [52]	96.97	85.79	72.01	84.23
ECT-net [53]	95.69	92.52	65.62	82.03
ConvNeXt-S [28]	91.09	78.11	91.85	81.94
CNN+Transformer [54]	95.83	91.72	80.33	85.33
CNN+LSTM [47]	96.42	95.31	89.95	92.50
CNN+Transformer [55]	95.55	88.13	70.65	87.55
ViT [29]	93.32	84.75	85.41	84.91
CNN-SVM [49]	97.53	91.24	85.63	91.00
CNN + BiLSTM [50]	97.39	96.56	90.11	94.20
PSC-Net (ours)	**98.85**	**97.01**	**96.86**	**97.33**

accuracy and improved generalizability compared to the evaluated baselines. The comparative experimental results are demonstrated in Tables 5 to 8. Based on the findings, it is evident that the PSC-Net model outperforms other models in the classification of the four datasets, namely, Five Types of Pulses Database, CHD Database, MIT-BIH Arrhythmia Database, and MIT-BIH ST Change Database. Our proposed method achieves impressive accuracies of 83.33%, 100.0%, 95.74%, and
98.85% for identifying the respective categories, along with F1-scores of 76.98%, 100.0%, 85.85%, and 97.33%, respectively. Meanwhile, other methods proposed by different researchers, which have also shown promising results in detecting pulse and cardiovascular diseases, still exhibit relatively lower F1-score and accuracy compared to the PSC-Net model. In Tables 5 to 8, CNN+BiLSTM ranks second in all indicators. This performance is attributed to the strengths of LSTM in handling long sequence data, preventing gradient vanishing problem, and long-term memory attributes that form the foundation of the GLNet constructed in this paper. Notably, standalone CNN architectures like ConvNeXt-S and standalone Transformer architecture ViT deliver only average performance. This is mainly due to their specialization in either local feature extraction or long-term feature extraction, which alone is insufficient for comprehensive model training. Additionally, traditional machine learning methods such as SVM and XGBoost exhibit subpar performance across all four evaluated metrics. This outcome is unsurprising, considering these methods are representative architectures widely used in the domain of impulse signals. This further emphasizes the remarkable contribution of our proposed approach to the advancement of wrist pulse signaling research.

The promising experimental outcomes in this paper can be largely attributed to the efficacy of the employed data augmentation algorithm. In particular, this data augmentation algorithm addresses the limitations of using the SMOTE Tomek algorithm alone for data augmentation. While this algorithm can generate additional samples, it may not ensure diverse feature representations. To enhance feature diversity, the designed data augmentation algorithm incorporates Gaussian noise into the SMOTE-fitted data. The augmented dataset is then merged with the original denoised training set. To ensure randomness, the Fisher–Yates algorithm is applied to randomize the dataset, resulting in a diverse training set suitable for model training. This strategy effectively mitigates issues related to data sparsity and class imbalance while promoting diversity in the generated data. This study selects the Five Types of Pulses Database and the Coronary Heart Disease (CHD) database as examples to demonstrate the application of data augmentation algorithms. Our purpose is to illustrate the process of applying the data augmentation algorithms to these specific datasets, highlighting their effectiveness in improving the quality of the data.

The specific distribution of the Five types of pulse database, CHD Database, MIT-BIH Arrhythmia Database and MIT-BIH ST Change Database after data augmentation are shown in Tables 9 to 12, respectively, in the form of three-line diagrams. The columns of the table exemplify the categories of the data, the number of samples in the test set, the number of samples in the training set, the total number of samples in the training set after data augmentation, the proportion of each category in the training set to the total data before data augmentation, and the proportion of each category in the training set to the total data after data augmentation, respectively. Analyzing the data in Tables 9 to 12 reveals the notable transformation in data distribution resulting from our data augmentation method. This underscored the efficacy of our proposed approach in mitigating data imbalance. For instance, Table 9 illustrates the impact of class imbalance on the training set, where the percentage of the Hesitant pulse is 3.26% and the percentage of the Moderate pulse is 45.74%. Such significant difference can adversely affect the overall experimental results. However, when employing the proposed data augmentation algorithm, each category accounts for approximately 15.01% of the total training set, effectively mitigating the data imbalance issue. In a similar manner, Table 10 presents the distribution of the Coronary Heart Disease (CHD) database after applying the data enrichment algorithm. In the original dataset, the CHD and the Normal categories account for 57.39% and 42.61% of the entire training set, respectively. However, after applying the data enrichment algorithm, both types are balanced at 53.44%. These results demonstrate the effectiveness of the designed data augmentation algorithm in addressing data sparsity and class imbalance, and promoting diversity within the training data.

5.4. Ablation experiments

In order to evaluate the effectiveness and individual contributions of the various modules in our proposed PSC-Net, a series of step-by-step ablation experiments was conducted on the Five Types of Pulses Database and MIT-BIH ST Change Database. The performance of the models was assessed by comparing them as follows:

Baseline: Our baseline model consists of GRWA.

Model 1: Based on the baseline, two layers of tandem LSTM are incorporated at the end of the model to create GLNet.

Model 2: Based on Model 1, a modified transformer encoder was incorporated, altering the two fully connected layers in the transform to GRU.

Model 3: Based on Model 2, residual connections containing global average pooling are added, connected from the output of the LSTM to the output of the Transformer encoder, and fused with a simple concatenation.

PSC-Net (ours): Based on Model 3, we replaced simple concatenation with an adjusted weight setting to enhance the integration of features.

Additionally, to compare the advantages and disadvantages of the constructed modules, the experimental results after replacing them with contemporary popular modules are arranged. Below are the models:

Model 4: Based on PSC-Net, the modified transformer encoder is replaced with the original transformer encoder.

Model 5: Based on PSC-Net, remove data augmentation operations.

Table 9
Five types of pulse database after data augmentation with the number of samples in each category.

Type of pulse	No. of beats	Test beats	Training beats	After data augmentation sampled beats (+SMOTE Tomek + Gaussian noise)	Percentage before data augmentation (%)	Percentage after data augmentation (%)
Hesitant pulse	99	24	75	1124	3.26	15.01
Slippery pulse	444	84	360	1395	15.65	18.63
Sluggish pulse	407	88	319	1358	13.86	18.14
Moderate pulse	1317	265	1052	2085	45.74	27.84
Counted pulse	609	115	494	1526	21.49	20.38
Total	2876	576	2300	7488	100.00	100.00

Table 10
CHD Database after data augmentation with the number of samples in each category.

Type of pulse	No. of beats	Test beats	Training beats	After data augmentation sampled beats (+SMOTE Tomek + Gaussian noise)	Percentage before data augmentation (%)	Percentage after data augmentation (%)
CHD	1270	303	967	1933	57.39	53.44
Normal	900	182	718	1684	42.61	46.56
Total	2170	485	1685	3617	100.0	100.0

Table 11
MIT-BIH Arrhythmia Database after data augmentation with the number of samples in each category.

Type of pulse	No. of beats	Test beats	Training beats	After data augmentation sampled beats (+SMOTE Tomek + Gaussian noise)	Percentage before data augmentation (%)	Percentage after data augmentation (%)
Normal beat	5921	1185	4737	9457	54.77	22.61
Left bundle branch block beat	949	189	760	5498	8.79	13.15
Right bundle branch block beat	977	195	782	5517	9.04	13.19
Atrial premature beat	540	108	432	5165	4.99	12.35
Premature ventricular contraction	2162	432	1730	6467	20.00	15.46
Fusion of ventricular and normal beat	217	43	174	4928	2.01	11.78
Fusion of paced and normal beat	43	9	34	4792	0.40	11.46
Total	10 809	2161	8649	41 824	100.00	100.00

Table 12
MIT-BIH ST Change Database after data augmentation with the number of samples in each category.

Type of pulse	No. of beats	Test beats	Training beats	After data augmentation sampled beats (+SMOTE Tomek + Gaussian noise)	Percentage before data augmentation (%)	Percentage after data augmentation (%)
Normal beat	6369	1274	5095	10 188	92.44	48.76
ST Elevation	478	96	328	5423	5.95	25.96
ST Depression	111	22	89	5279	1.61	25.28
Total	6958	1392	5512	20 890	100.0	100.0

Table 13
Comparison of performance on the Five Types of Pulses Database and MIT-BIH ST Change Database after the gradual addition of each module of the model. The best results are shown in bold.

Models	Database metric	Five Types of Pulses Database				MIT-BIH ST Change Database			
		ACC (%)	PR (%)	RE (%)	F1 (%)	ACC (%)	PR (%)	RE (%)	F1 (%)
Baseline		71.18	64.93	66.29	65.27	94.40	87.51	70.64	77.32
Model 1		78.82	70.82	68.86	69.28	96.41	85.70	88.65	97.12
Model 2		77.04	74.54	69.88	71.24	97.13	91.89	89.79	90.37
Model 3		79.86	69.80	68.21	68.73	97.84	92.46	94.08	93.19
PSC-Net (ours)		**83.33**	**79.93**	**75.52**	**76.99**	**98.85**	**97.01**	**96.86**	**97.33**

Table 14
Presents a comparative analysis of the performance when each component of the model is replaced with modules from previous studies. The best results are shown in bold.

Models	Database metric	Five Types of Pulses Database				MIT-BIH ST Change Database			
		ACC (%)	PR (%)	RE (%)	F1 (%)	ACC (%)	PR (%)	RE (%)	F1 (%)
Model 4		80.21	75.80	72.08	73.33	97.41	88.97	87.32	88.12
Model 5		79.51	72.05	69.43	69.48	97.13	93.58	83.60	87.86
Model 6 [27]		79.92	73.62	71.55	72.69	97.59	94.59	84.96	87.56
Model 7 [56]		81.68	73.66	72.29	74.38	98.12	95.64	93.96	90.35
Model 8 [57]		**82.91**	76.98	75.42	76.55	98.10	97.33	97.41	**96.02**
Model 9		75.00	75.07	66.23	67.87	**98.71**	**97.56**	85.05	89.95
Model 10		72.18	63.08	63.10	62.98	98.42	93.82	93.24	93.52
Model 11		75.35	69.38	68.34	68.47	98.71	93.77	95.03	94.05
Model 12		79.51	73.55	75.59	74.48	97.85	96.07	94.68	95.80
Model 13		81.94	**78.76**	**76.37**	**76.98**	98.71	93.73	**97.60**	95.59

Model 6: Based on Model 6, SMOTE was employed as the data augmentation algorithm.

Model 7: Based on Model 6, BAGAN was employed as the data augmentation algorithm.

Model 8: Based on Model 6, GAMO was employed as the data augmentation algorithm.

Model 9: Based on PSC-Net, the loss function will be modified to a single loss function, namely, cross-entropy.

Model 10: Based on PSC-Net, the loss function will be modified to a single loss function, specifically focal loss.

Model 11: Based on Model 11, the automatic weight settings proposed in this paper are added.

Model 12: Based on PSC-Net, remove the optimization of DBA.

Model 13: Based on Model 12, an annealed cosine optimization algorithm is added to optimize the model.

We systematically evaluated each module and reported the outcomes in Tables 13 and 14. Table 13 shows the performance of each part of the proposed PSC-Net in two datasets to demonstrate the effectiveness of each part. After that, Table 14 compares the experimental results of the model components after replacing them with the modules from the existing study.

The results unambiguously indicate that our PSC-Net outperforms the individual CNN and Transformer modules in terms of both F1-score and accuracy. This superiority can be attributed to the proficiency of CNN in capturing local features and the excellence of the Transformer in extracting global features. By integrating both modules, our model achieves a comprehensive representation of intracycle and intercycle features alike. These outcomes underscore the superior learning capacity of the combined intra-periodic and inter-periodic features for detecting physiological signal feature waveforms. At the same time, by employing feature fusion together with optimized weights instead of direct concatenation, notable improvements in accuracy were observed on both the Five Types of Pulses Database and the MIT-BIH ST Change Database. Specifically, the accuracy on the Five Types of Pulses Database increased by 3.47%, while on the MIT-BIH ST Change Database, the accuracy improved by 1.01%. Moreover, the F1-score, which measures the balance between precision and recall, also exhibited significant enhancements. On the Five Types of Pulses Database, the F1-score improved by 8.78%, and on the MIT-BIH ST Change Database, the F1-score increased by 4.14%. The advantage of setting weights for feature fusion over direct concatenation lies in its ability to assign different levels of importance to individual features during the fusion process. By adjusting the weights, more emphasis can be placed on informative features while reducing the influence of less discriminative ones. This strategy allows for a more fine-grained integration of features, capturing their complementary nature and enhancing the overall discriminative power of fused representation.

In addition, the proposed data augmentation algorithm effectively mitigates the adverse effects of data scarcity and data imbalance on model accuracy. As depicted in Fig. 4(c), the augmented dataset demonstrates improved data volume and a more balanced category distribution compared to Fig. 4(a). As can be seen from Table 14, our model achieves a significant accuracy improvement of 3.82% and 1.29% for the respective metrics compared to the scenario without data augmentation.

A further notable contribution of this paper lies in the design of an effective loss function. By assigning appropriate weights to the loss function, the benefits of the three individual loss functions are combined, leading to significant improvements in various evaluation metrics compared to using a single loss function. Importantly, for the weight setting of the focal loss, an algorithm is devised to automatically determine the optimal weight. This automated approach surpasses manual weight setting and achieves a remarkable accuracy improvement of 3.2% and F1-score improvement of 5.49% on the Five Types of Pulses Database. In a similar manner, on the MIT-BIH ST Change Database, the accuracy is enhanced by 0.29% and the F1-score shows a notable increase of 0.53%. This is mainly due to the fact that the automatic weight assignment in focal loss improves adaptation by assigning higher weights to challenging samples or classes, enabling the model to prioritize difficult examples and enhance the performance on unbalanced datasets; it specifically enhances learning for minority groups and mitigates the impact of class imbalance.

Lastly, we used the global search capability of the DBA to accelerate the identification of optimal learning rate for the ongoing training phase, thereby increasing the convergence rate of the model. Table 14 presents the results of the ablation experiments, comparing the performance of the DBA proposed in this study with the widely used annealing cosine algorithm. The results demonstrate that the DBA outperforms the annealing cosine algorithm in terms of accuracy. Specifically, on the Five Types of Pulses Database, it achieves an accuracy improvement of 1.39% compared to the annealing cosine algorithm. In a similar manner, on the MIT-BIH ST Change Database, the DBA shows a performance improvement of 0.14% over the annealing cosine algorithm. It is worth noting that the algorithm without any optimization achieves lower accuracy, with 79.51% for Five Types of Pulses Database and 97.85% for MIT-BIH ST Change Database. These findings highlight the effectiveness of the DBA in enhancing the accuracy of the classification task when compared to existing optimization approaches. In contrast, the accuracy of the algorithm without any optimization is only 79.51% for the Five Types of Pulses Database and 97.85% for the MIT-BIH ST Change Database. It is noteworthy that the adaptability of the DBA allows it to dynamically adjust the learning rate according to the dynamic changes in the training process, alleviating the oscillations and instabilities often encountered during training and ultimately improving the generalization of the model. This is one of the reasons for the smooth curves seen in Fig. 11(a) and (b).

Fig. 11(c) demonstrates that despite the evident disparities in category sizes within the original dataset, our data processing approach ensures that the significant differences in AUC values between categories are minimized. Remarkably, the majority of categories achieve AUC values exceeding 90%. However, it is noteworthy that category 0 still exhibits relatively weaker AUC values compared to other categories, primarily due to its substantial disparity in category size. Resolving this disparity will be a focal point for future investigations.

6. Discussion

Our model, designed specifically for waveform detection in physiological signals, surpasses most existing algorithms in real-world applicability. It effectively addresses challenges related to data imbalance and sparsity, enabling accurate identification of small sample data using deep learning models. This aspect is particularly advantageous for disease datasets characterized by data scarcity and imbalance, commonly encountered issues in the medical field. Moreover, our model incorporates crucial features of physiological signals by leveraging modules such as LSTM and GRU to capture data trends efficiently. The fused architecture of CNNs and Transformers can extract valuable feature variables within and between cycles. The specialized one-dimensional residual blocks designed for physiological signals play a pivotal role in achieving exceptional performance, especially in frequency domain feature extraction.

In the GRWA block, our proposed methodology harnesses a synergistic blend of causal convolution, dilated convolution, and gated convolution. Causal convolution is strategically employed to adeptly consider the causal relationship within the temporal domain, while dilated convolution expands the receptive field and extracts a richer array of salient features, as vividly illustrated in Fig. 7. Incorporate gated convolution regulates the flow of information via gating mechanisms, empowering our neural network to capture intricate long-term dependencies inherent in the input signal.

Two LSTM layers were added to form the GLNet model architecture. According to the results in Tables 5, 8 and 14, compared to

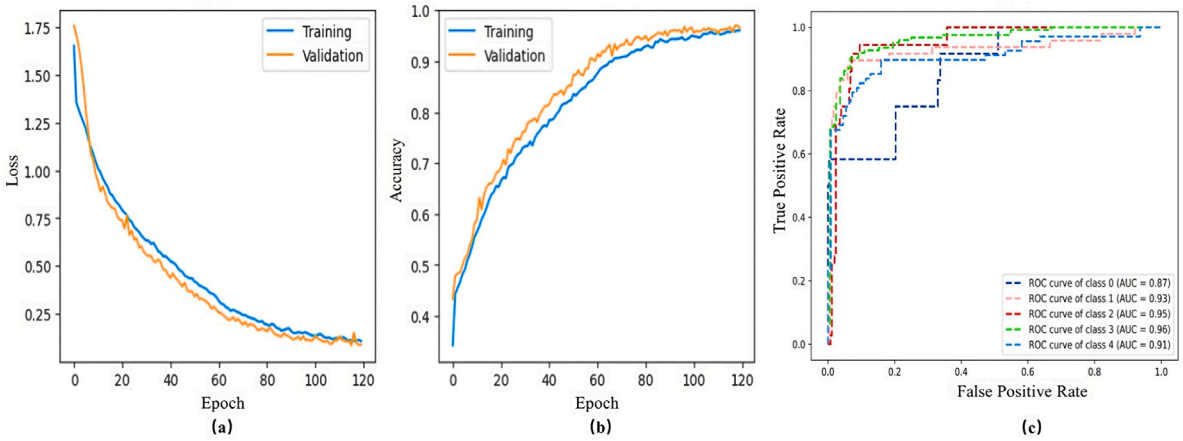


Fig. 11. The Five Types of Pulses Database is selected as a sample: (a) The accuracy loss function curves for the training and validation sets; (b) The accuracy curves for the training and validation sets; (c) The class AUC curves for the test set, classes 0, 1, 2, 3, 4 correspond to Hesitant pulse, Slippery pulse, Sluggish pulse, Moderate pulse, Counted pulse.

the ConvNeXt-S architecture, the GLNet architecture shows a 1.05% improvement in accuracy, a 6.03% improvement in precision, a 5% improvement in recall, and a 5.23% improvement in F1 scores in the Five Types of Pulses Database. Additionally, the GLNet architecture outperforming the ConvNeXt-S architecture by 5.32%, 7.59%, and 15.18% on the three evaluated metrics (accuracy, precision, and F1 score, respectively) on the publicly available MIT-BIH ST Change Database. Notably, ConvNeXt-S excels in negative sample detection on the MIT-BIH ST Change Database.

To address, the CNN's limited capability for capturing long-range dependencies [58], we integrated a Transformer encoder architecture. Unlike the conventional parallelization of CNN and Transformer structures for feature extraction, we introduced GLNet to preprocess data, allowing us to extract long-range feature dependencies while preserving local features through a residual connection. Experimental results in Table 14 for Model 2 and PSC-Net on the Five Types of Pulses Database demonstrate a 6.29% increase in accuracy and a 5.75% boost in F1 score, showcasing the effectiveness of our novel architecture. At the same time, we replaced the Dense layer in the original Transformer encoder with a GRU architecture, better suited for time-series data processing. Comparative results in Table 14 reveal that in the Five Types of Pulses Database, PSC-Net outperforms Model 4 by 3.12% in accuracy, 4.13% in precision, 3.44% in recall, and 3.66% in F1-score. On the publicly available MIT-BIH ST Change Database, these four metrics exhibit improvements of 1.44%, 8.04%, 9.54%, and 9.21%.

The main reasons for choosing to use SMOTE Tomek instead of GAN, diffusion models and other methods in data augmentation are:

- The dataset in this paper is one-dimensional data, and GAN, diffusion models are more often used in images. One-dimensional data feature points are relatively simple and do not require complex data generation techniques.
- GAN models usually need a lot of data for support, and this condition is difficult to achieve in the field of medical diagnosis.
- SMOTE Tomek is easy to use and can be interpreted. Its implementation principle is to use the SMOTE algorithm to synthesize a new minority class of samples and then use the TOMEK algorithm to remove the synthesized samples as noise. The generated data can be directly visualized using t-sne, PCA and other methods, as shown in Fig. 4.

The experiments in this study confirm the effectiveness of the proposed data enhancement method. Table 13 illustrates improved results for PSC-Net on both datasets. Conversely, the model without data enhancement, Model 5, exhibits a notable performance decline. More precisely, within the Five Types of Pulses Database, accuracy, precision, recall, and F1-score decreased by 3.82%, 7.88%, 6.09%, and 7.51%, respectively. Likewise, on the MIT-BIH ST Change Database, the respective evaluation metrics experienced decreases of 1.72%, 3.43%, 13.26%, and 9.47%. These findings underscore the positive influence of the proposed data enhancement approach on the model's performance.

In addition, the data augmentation method used in this paper is improved based on the SMOTE Tomek algorithm. We found the optimal data augmentation method by replacing the SMOTE Tomek in this paper with SMOTE, BAGAN, and GAMO, respectively. In the related experiments conducted for Five types of pulse database, the results in Table 14 show that the data augmentation algorithm we used based on the SMOTE Tomek improvement achieved the best performance. Specifically, it outperforms SMOTE by 4.3%, BAGAN by 2.61% and GAMO by 0.44% in terms of accuracy. For the field of data augmentation, our next work will focus on combining SMOTE and GAN to explore whether we can improve a new algorithm based on the advantages of both.

Precise recognition of physiological signals can aid doctors in diagnosing physiological diseases, playing a crucial role in early diagnosis and prevention. Extensive testing and analysis across diverse physiological signal datasets encompassing various diseases consistently demonstrate the superior performance of our proposed approach. These findings validate the practical value of our network in supporting clinical medical diagnosis. Notably, physiological signal diagnosis often requires considering multiple conditions. While our current focus is on the time and frequency domains of the dataset, future investigations will delve into other domains to provide more comprehensive insights. Moreover, our future work will involve integrating temporal information with knowledge graphs to provide a more usable method for physiological signal recognition.

7. Conclusions

The proposed PSC-Net model architecture aims to efficiently process physiological signal data for accurate disease diagnosis. This model takes patient data as input and outputs corresponding categories, providing doctors with a faster and more accurate means of assessing the patient's physical condition. Physiologic Signal Diagnostics, being non-invasive and reproducible, offers a valuable tool for continuous monitoring and early detection of health issues. Common healthcare devices can readily acquire relevant medical data suitable for deep learning analysis. PSC-Net combines Convolutional Neural Networks (CNNs) and Transformers, effectively utilizing morphological and temporal information in physiological signals for enhanced detection performance.

Specifically, GLNet generates ripple embedding features, facilitating the efficient extraction of intra-cycle features without excessive computational complexity. The global average pooling in residual concatenation aids in sampling frequency domain features in continuous physiological signals. The residual linking fosters interaction between CNN and Transformer encoder, filtering features for the Transformer input without neglecting local features or discarding redundant information. Furthermore, the inclusion of data augmentation, focal loss automatic weight setting, and the use of the DBA contribute to enhancing the accuracy of the model.

Extensive evaluation of two pulse datasets demonstrated the superiority, PSC-Net over existing models, achieving excellent performance with 83.33% accuracy, 79.93% precision, 75.52% recall, and 76.99% F1-score across Five Types of Pulses Database. The CHD Database exhibited outstanding detection capabilities, showcasing a perfect 100.0% performance across all four evaluation metrics. Moreover, our method consistently performed well on two ECG public datasets, indicating its stability and generalization ability. Our proposed PSC-Net has proved to be a promising diagnostic tool in clinical applications for accurately detecting characteristic waveforms in interpatient pathology data, and thus to derive the presence or absence of physiological signaling abnormalities, and to provide an aid to physician disease diagnosis. The accompanying proposed data augmentation algorithm and weight autosetting algorithm hold great potential in diverse applications in the relevant field. Future work is aimed at optimizing the detection accuracy and exploring disease diagnosis using additional domain features.

The manuscript is confirmed to be exclusively submitted to this journal and is not under consideration or published elsewhere. The authors affirm that there are no known financial or personal relationships that could have affected the findings presented in this article. No potential conflicts of interest exist related to the research, authorship, or publication of this work.

CRediT authorship contribution statement

Qichao Liu: Methodology, Visualization, Writing – original draft, Writing – review & editing. **Yue Feng:** Project administration, Resources, Supervision, Writing – review & editing. **Hong Xu:** Funding acquisition, Writing – review & editing. **Jia Li:** Writing – review & editing. **Zhuosheng Lin:** Writing – review & editing. **Shengke Li:** Writing – review & editing. **Shihan Qiu:** Writing – review & editing. **Xin Wu:** Writing – review & editing. **Yuangang Ma:** Writing – review & editing. **Ying Xu:** Data curation. **Fufeng Li:** Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work is supported by the Basic Research and Applied Basic Research Key Project in General Colleges and Universities of Guangdong Province, China (2021ZDZX1032); the Special Project of Guangdong Province, China (2020A13 13030021); and the Scientific Research Project of Wuyi University (2018TP023, 2018GR003). Qichao Liu and Yue Feng contributed equally to this article.

References

[1] Adam Timmis, Panos Vardas, Nick Townsend, Aleksandra Torbica, Hugo Katus, Delphine De Smedt, Chris P Gale, Aldo P Maggioni, Steffen E Petersen, Radu Huculeci, et al., European society of cardiology: cardiovascular disease statistics 2021, Eur. Heart J. 43 (8) (2022) 716–799, http://dx.doi.org/10.1093/eurheartj/ehab892.

[2] Chaoxun Guo, Zhixing Jiang, Haoze He, Yining Liao, David Zhang, Wrist pulse signal acquisition and analysis for disease diagnosis: A review, Comput. Biol. Med. 143 (2022) 105312, http://dx.doi.org/10.1016/j.compbiomed.2022.105312.

[3] Qi Zhang, Jianhang Zhou, Bob Zhang, Graph based multichannel feature fusion for wrist pulse diagnosis, IEEE J. Biomed. Health Inf. 25 (10) (2020) 3732–3743, http://dx.doi.org/10.1109/JBHI.2020.3045274.

[4] David Cyranoski, Why Chinese medicine is heading for clinics around the world, Nature 561 (7724) (2018) 448–448, http://dx.doi.org/10.1038/d41586-018-06782-7.

[5] Sajad Mousavi, Fatemeh Afghah, Fatemeh Khadem, U Rajendra Acharya, ECG language processing (ELP): A new technique to analyze ECG signals, Comput. Methods Programs Biomed. 202 (2021) 105959, http://dx.doi.org/10.1016/j.cmpb.2021.105959.

[6] Li Guo, Gavin Sim, Bogdan Matuszewski, Inter-patient ECG classification with convolutional and recurrent neural networks, Biocybern. Biomed. Eng. 39 (3) (2019) 868–879, http://dx.doi.org/10.1016/j.bbe.2019.06.001.

[7] Yuanlu Li, Renfei Qian, Kun Li, Inter-patient arrhythmia classification with improved deep residual convolutional neural network, Comput. Methods Programs Biomed. 214 (2022) 106582, http://dx.doi.org/10.1016/j.cmpb.2021.106582.

[8] Dimin Wang, David Zhang, Guangming Lu, A robust signal preprocessing framework for wrist pulse analysis, Biomed. Signal Process. Control 23 (2016) 62–75, http://dx.doi.org/10.1016/j.bspc.2015.08.002.

[9] Manasvi Kashyap, Shruti Jain, Importance of pulse examination and its diagnostic system, in: Recent Innovations in Computing: Proceedings of ICRIC 2021, Volume 1, Springer, 2022, pp. 189–199, http://dx.doi.org/10.1007/978-981-16-8248-3_16.

[10] Jian Cui, Lucheng Song, Wrist pulse diagnosis of stable coronary heart disease based on acoustics waveforms, Comput. Methods Programs Biomed. 214 (2022) 106550, http://dx.doi.org/10.1016/j.cmpb.2021.106550.

[11] Karthikeyan Venugopal, Parthasarathy Panchatcharam, Arunkumar Chandrasekhar, Vivekanandan Shanmugasundaram, Comprehensive review on triboelectric nanogenerator based wrist pulse measurement: Sensor fabrication and diagnosis of arterial pressure, ACS Sens. 6 (5) (2021) 1681–1694, http://dx.doi.org/10.1021/acssensors.0c02324.

[12] Chuanglu Chen, Zhiqiang Li, Yitao Zhang, Shaolong Zhang, Jiena Hou, Haiying Zhang, A 3D wrist pulse signal acquisition system for width information of pulse wave, Sensors 20 (1) (2019) 11, http://dx.doi.org/10.3390/s20010011.

[13] Xiaoxiao Kang, Jun Zhang, Zheming Shao, Guotai Wang, Xingguang Geng, Yitao Zhang, Haiying Zhang, A wearable and real-time pulse wave monitoring system based on a flexible compound sensor, Biosensors 12 (2) (2022) 133, http://dx.doi.org/10.3390/bios12020133.

[14] Dimin Wang, David Zhang, Guangming Lu, A novel multichannel wrist pulse system with different sensor arrays, IEEE Trans. Instrum. Meas. 64 (7) (2015) 2020–2034, http://dx.doi.org/10.1109/TIM.2014.2357599.

[15] Ahmad Almadhor, Gabriel Avelino Sampedro, Mideth Abisado, Sidra Abbas, Efficient feature-selection-based stacking model for stress detection based on chest electrodermal activity, Sensors 23 (15) (2023) 6664, http://dx.doi.org/10.3390/s23156664.

[16] Tom Ganz, Erik Imgrund, Martin Härterich, Konrad Rieck, CodeGraphSMOTE-data augmentation for vulnerability discovery, in: IFIP Annual Conference on Data and Applications Security and Privacy, Springer, 2023, pp. 282–301, http://dx.doi.org/10.1007/978-3-031-37586-6_17.

[17] Yongxu Liu, Yan Liu, XB Bruce, Shenghua Zhong, Zhejing Hu, Noise-robust oversampling for imbalanced data classification, Pattern Recognit. 133 (2023) 109008, http://dx.doi.org/10.1016/j.patcog.2022.109008.

[18] Shweta Sharma, Anjana Gosain, Shreya Jain, A review of the oversampling techniques in class imbalance problem, in: International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 1, Springer, 2022, pp. 459–472, http://dx.doi.org/10.1007/978-981-16-2594-7_38.

[19] E Smily JeyaJothi, J Anitha, Shalli Rani, Basant Tiwari, et al., A comprehensive review: computational models for obstructive sleep apnea detection in biomedical applications, BioMed Res. Int. 2022 (2022) http://dx.doi.org/10.1155/2022/7242667.

[20] E. Quanyu, Pulse signal analysis based on deep learning network, BioMed Res. Int. 2022 (2022) http://dx.doi.org/10.1155/2022/6256126.

[21] Nan Li, Jiarui Yu, Xiaobo Mao, Yuping Zhao, Luqi Huang, A hybrid 1DCNN-GRU-attention model for three-stage pregnancy pulse identification, 2023, http://dx.doi.org/10.21203/rs.3.rs-2314703/v1.

[22] G.C. Suguna, S.T. Veerabhadrappa, Denoising wrist pulse signals using variance thresholding technique, Indian J. Sci. Technol. 13 (41) (2020) 4275–4286, http://dx.doi.org/10.17485/IJST/v13i40.1625.

[23] Sakib Mahmud, Nabil Ibtehaz, Amith Khandakar, M Sohel Rahman, Antonio JR Gonzales, Tawsifur Rahman, Md Shafayet Hossain, Md Sakib Abrar Hossain, Md Ahasan Atick Faisal, Farhan Fuad Abir, et al., NABNet: a nested attention-guided BiConvLSTM network for a robust prediction of blood pressure components from reconstructed arterial blood pressure waveforms using PPG and ECG signals, Biomed. Signal Process. Control 79 (2023) 104247, http://dx.doi.org/10.1016/j.bspc.2022.104247.

[24] Damien Dablain, Bartosz Krawczyk, Nitesh V. Chawla, DeepSMOTE: Fusing deep learning and SMOTE for imbalanced data, IEEE Trans. Neural Netw. Learn. Syst. 34 (9) (2023) 6390–6404, http://dx.doi.org/10.1109/TNNLS.2021.3136503.

[25] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron Courville, Improved Training of Wasserstein GANs Montreal Institute for Learning Algorithms, Technical Report, Technical Report, 2020, http://dx.doi.org/10.48550/arXiv.1704.00028, arXiv 2017.

[26] Kechen Shu, Shitong Mao, James L. Coyle, Ervin Sejdić, Improving non-invasive aspiration detection with auxiliary classifier wasserstein generative adversarial networks, IEEE J. Biomed. Health Inform. 26 (3) (2021) 1263–1272, http://dx.doi.org/10.1109/JBHI.2021.3106565.

[27] Fadel M Megahed, Ying-Ju Chen, Aly Megahed, Yuya Ong, Naomi Altman, Martin Krzywinski, The class imbalance problem, Nature Methods 18 (11) (2021) 1270–1272.

[28] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining Xie, A convnet for the 2020s, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976–11986, http://dx.doi.org/10.48550/arXiv.2201.03545.

[29] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, http://dx.doi.org/10.48550/arXiv.2010.11929, arXiv preprint arXiv:2010.11929.

[30] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022, http://dx.doi.org/10.48550/arXiv.2103.14030.

[31] Haixu Wu, Jiehui Xu, Jianmin Wang, Mingsheng Long, Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, Adv. Neural Inf. Process. Syst. 34 (2021) 22419–22430, http://dx.doi.org/10.48550/arXiv.2106.13008.

[32] George B. Moody, Roger G. Mark, The impact of the MIT-BIH arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50, http://dx.doi.org/10.1109/51.932724.

[33] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, H Eugene Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220, http://dx.doi.org/10.1161/01.cir.101.23.e215.

[34] Li Gong, Zhiqing Xiao, Lihua Xu, Yan Ding, Zhuo Zou, Lirong Zheng, An IoT-based wearable labor progress monitoring system for remote evaluation of admission time to hospital, IEEE J. Biomed. Health Inf. (2023) http://dx.doi.org/10.1161/01.cir.101.23.e215.

[35] Yanqin Yang, Xinge Guo, Minglu Zhu, Zhongda Sun, Zixuan Zhang, Tianyiyi He, Chengkuo Lee, Triboelectric nanogenerator enabled wearable sensors and electronics for sustainable internet of things integrated green earth, Adv. Energy Mater. 13 (1) (2023) 2203040, http://dx.doi.org/10.1002/aenm.202203040.

[36] Byeonghak Park, Joo Hwan Shin, Jehyung Ok, Subin Park, Woojin Jung, Chanho Jeong, Seunghwan Choy, Young Jin Jo, Tae-il Kim, Cuticular pad–inspired selective frequency damper for nearly dynamic noise–free bioelectronics, Science 376 (6593) (2022) 624–629, http://dx.doi.org/10.1126/science.abj9912.

[37] Shahid A. Malik, Shabir A. Parah, Bilal A. Malik, Power line noise and baseline wander removal from ECG signals using empirical mode decomposition and lifting wavelet transform technique, Health Technol. 12 (4) (2022) 745–756, http://dx.doi.org/10.1007/s12553-022-00662-x.

[38] Jie Wang, Yirun Zhu, Zhiyong Wu, Yunlin Zhang, Jian Lin, Tao Chen, Huicong Liu, Fengxia Wang, Lining Sun, Wearable multichannel pulse condition monitoring system based on flexible pressure sensor arrays, Microsyst. Nanoeng. 8 (1) (2022) 16, http://dx.doi.org/10.1038/s41378-022-00349-3.

[39] Georgios Petmezas, Kostas Haris, Leandros Stefanopoulos, Vassilis Kilintzis, Andreas Tzavelis, John A Rogers, Aggelos K Katsaggelos, Nicos Maglaveras, Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets, Biomed. Signal Process. Control 63 (2021) 102194, http://dx.doi.org/10.1016/j.bspc.2020.102194.

[40] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017) http://dx.doi.org/10.48550/arXiv.1706.03762.

[41] Jun Chen, HongSheng Mai, Linbo Luo, Xiaoqiang Chen, Kangle Wu, Effective feature fusion network in BIFPN for small object detection, in: 2021 IEEE International Conference on Image Processing, ICIP, IEEE, 2021, pp. 699–703, http://dx.doi.org/10.1109/ICIP42928.2021.9506347.

[42] Jing Zhang, Aiping Liu, Min Gao, Xiang Chen, Xu Zhang, Xun Chen, ECG-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network, Artif. Intell. Med. 106 (2020) 101856, http://dx.doi.org/10.1016/j.artmed.2020.101856.

[43] Ghayth AlMahadin, Ahmad Lotfi, Marie Mc Carthy, Philip Breedon, Enhanced Parkinson's disease tremor severity classification by combining signal processing with resampling techniques, SN Comput. Sci. 3 (2022) 1–21, http://dx.doi.org/10.1007/s42979-021-00953-6.

[44] Julien Fleuret, Samira Ebrahimi, C Ibarra Castanedo, X Maldague, On the use of pulsed thermography signal reconstruction based on linear support vector regression for carbon fiber reinforced polymer inspection, Quant. InfraRed Thermogr. J. 20 (2) (2023) 39–61, http://dx.doi.org/10.1080/17686733.2021.2025015.

[45] Hsing-Chung Chen, Cahya Damarjati, Karisma Trinanda Putra, Han-MI Chen, Ching-Liang Hsieh, Hung-Jen Lin, Mei-Yao Wu, Chin-Sheng Chen, Pulse-line intersection method with unboxed artificial intelligence for hesitant pulse wave classification, Inf. Process. Manage. 59 (2) (2022) 102855, http://dx.doi.org/10.1016/j.ipm.2021.102855.

[46] Zhang-Meng Liu, S. Yu Philip, Classification, denoising, and deinterleaving of pulse streams with recurrent neural networks, IEEE Trans. Aerosp. Electron. Syst. 55 (4) (2018) 1624–1639, http://dx.doi.org/10.1109/TAES.2018.2874139.

[47] Runnan He, Yang Liu, Kuanquan Wang, Na Zhao, Yongfeng Yuan, Qince Li, Henggui Zhang, Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional LSTM, IEEE Access 7 (2019) 102119–102135, http://dx.doi.org/10.1109/ACCESS.2019.2931500.

[48] Jianhong Chen, Huang Huang, Wenrui Hao, Jinchao Xu, A machine learning method correlating pulse pressure wave data with pregnancy, Int. J. Numer. Methods Biomed. Eng. 36 (1) (2020) e3272, http://dx.doi.org/10.1002/cnm.3272.

[49] Manoj Kumar Ojha, Sulochna Wadhwani, Arun Kumar Wadhwani, Anupam Shukla, Automatic detection of arrhythmias from an ECG signal using an auto-encoder and SVM classifier, Phys. Eng. Sci. Med. 45 (2) (2022) 665–674, http://dx.doi.org/10.1007/s13246-022-01119-1.

[50] Wissal Midani, Wael Ouarda, Mounir Ben Ayed, DeepArr: An investigative tool for arrhythmia detection using a contextual deep neural network from electrocardiograms (ECG) signals, Biomed. Signal Process. Control 85 (2023) 104954, http://dx.doi.org/10.1016/j.bspc.2023.104954.

[51] Adel A Ahmed, Waleed Ali, Talal AA Abdullah, Sharaf J Malebary, Classifying cardiac arrhythmia from ECG signal using 1D CNN deep learning model, Mathematics 11 (3) (2023) 562, http://dx.doi.org/10.3390/math11030562.

[52] Taotao Liu, Yujuan Si, Weiyi Yang, Jiaqi Huang, Yongheng Yu, Gengbo Zhang, Rongrong Zhou, Inter-patient congestive heart failure detection using ECG-convolution-vision transformer network, Sensors 22 (9) (2022) 3283, http://dx.doi.org/10.3390/s22093283.

[53] Duoduo Wang, Lishen Qiu, Wenliang Zhu, Yanfang Dong, Huimin Zhang, Yuhang Chen, et al., Inter-patient ECG characteristic wave detection based on convolutional neural network combined with transformer, Biomed. Signal Process. Control 81 (2023) 104436, http://dx.doi.org/10.1016/j.bspc.2022.104436.

[54] Chao Che, Peiliang Zhang, Min Zhu, Yue Qu, Bo Jin, Constrained transformer network for ECG signal processing and arrhythmia classification, BMC Med. Inform. Decis. Mak. 21 (1) (2021) 1–13, http://dx.doi.org/10.1186/s12911-021-01546-2.

[55] Behnam Behinaein, Anubhav Bhatti, Dirk Rodenburg, Paul Hungler, Ali Etemad, A transformer architecture for stress detection from ecg, in: Proceedings of the 2021 ACM International Symposium on Wearable Computers, 2021, pp. 132–134, http://dx.doi.org/10.48550/arXiv.2108.09737.

[56] Giovanni Mariani, Florian Scheidegger, Roxana Istrate, Costas Bekas, Cristiano Malossi, Bagan: Data augmentation with balancing gan, 2018, http://dx.doi.org/10.48550/arXiv.1803.09655, arXiv preprint arXiv:1803.09655.

[57] Swalpa Kumar Roy, Juan M Haut, Mercedes E Paoletti, Shiv Ram Dubey, Antonio Plaza, Generative adversarial minority oversampling for spectral–spatial hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–15, http://dx.doi.org/10.1109/TGRS.2021.3052048.

[58] Changqing Ji, Liyong Wang, Jing Qin, Lu Liu, Yue Han, Zumin Wang, MS-Gformer: A multi-scale grid transformer network for 12-lead ECG arrhythmia detection, Biomed. Signal Process. Control 87 (2024) 105499, http://dx.doi.org/10.1016/j.bspc.2023.105499.